

Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions

Year: 2023 | Citations: 316 | Authors: Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G. Parker, M. de Choudhury

Abstract

Large language models have abilities in creating high-volume human-like texts and can be used to generate persuasive misinformation. However, the risks remain under-explored. To address the gap, this work first examined characteristics of AI-generated misinformation (AI-misinfo) compared with human creations, and then evaluated the applicability of existing solutions. We compiled human-created COVID-19 misinformation and abstracted it into narrative prompts for a language model to output AI-misinfo. We found significant linguistic differences within human-AI pairs, and patterns of AI-misinfo in enhancing details, communicating uncertainties, drawing conclusions, and simulating personal tones. While existing models remained capable of classifying AI-misinfo, a significant performance drop compared to human-misinfo was observed. Results suggested that existing information assessment guidelines had questionable applicability, as AI-misinfo tended to meet criteria in evidence credibility, source transparency, and limitation acknowledgment. We discuss implications for practitioners, researchers, and journalists, as AI can create new challenges to the societal problem of misinformation.